In presenting the dissertation as a partial fulfillment of the requirements for an advanced degree from the Georgia Institute of Technology, I agree that the Library of the Institute shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to copy from, or to publish from, this dissertation may be granted by the professor under whose direction it was written, or, in his absence, by the Dean of the Graduate Division when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from, or publication of, this dissertation which involves potential financial gain will not be allowed without written permission.

7/25/68

ERRATUM

Page 64 is missing; however, this is an error in pagination and does not represent an error in the text.

APPLICATIONS OF REPRODUCING KERNELS IN HILBERT SPACES

A THESIS

Presented to

The Faculty of the Division of Graduate

Studies and Research

by

Michael Leslie Mumford

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Applied Mathematics

Georgia Institute of Technology

April, 1972

APPLICATIONS OF REPRODUCING KERNELS IN HILBERT SPACES

Approved:

Chairman

Date approved by Chairman: 5/10/72

DEDICATION

To my wife,

Jerry,

who has great love and great patience.

ACKNOWLEDGMENTS

I wish to thank my advisor, Dr. W. J. Kammerer, for his assistance and encouragement during the preparation of this thesis. I would also like to thank the members of my reading committee, Dr. J. V. Herod and Dr. R. H. Kasriel, for their valuable suggestions.

The initial work on this thesis was begun under a grant from the National Science Foundation in the summer of 1971. This assistance is gratefully acknowledged.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

CHAPTER I

INTRODUCTION

The purpose of this work is to explore some applications of the theory of reproducing kernels and to examine some of the connections between historically different approaches. The basis of the theory of reproducing kernels was laid down by N. Aronszajn [1] in 1950, and for our purposes the relevant aspects of his work are presented in the introductory theory in this chapter.

Chapter II begins with the problem of interpolation at a finite number of points, and follows the ideas of B. Chalmers [4] in an extension to an infinite number of points or functionals. From K. Yao [17] the idea of a sampling expansion and the space H_1 are introduced in Section Two, with some considerations of errors in finite approximations to infinite sums. Section Three explores the theory of splines and examines several minimum properties with the use of the reproducing kernel. The last section in this chapter contains an application to iterative processes.

The approximation properties of the first section of Chapter III are suggested by work with the spline functions of Chapter II, after some work by C. de Boor and R.E.

Lynch [3]; additionally, Section Two examines approximation in the sense of Sard in spaces with spline functions. Section Three reexamines some approximation problems using a constructive technique which directly gives the approximation constants.

Let X be a Hilbert space of functions defined on a set D, the inner product of f with g being denoted by (f,g).

Definition 1.1. A reproducing kernel for X is a function from DxD to the scalar field of X which has the following properties:

(i) For any fixed y in D, $k(\cdot,y)$ is an element of X

(ii) For any f in X, y in D, $f(y) = (f(x),k(x,y))$.

It is nice not to have too many objects around in order to avoid confusion, and thus the following is true:

Theorem 1.1. If X is a Hilbert space with a reproducing kernel $k(x,y)$, then the kernel is unique.

Proof. Suppose $k(x,y)$, $j(x,y)$ are each reproducing kernels for X. Then for fixed y,

$$||k(\cdot,y)-j(\cdot,y)||^2 = (k(x,y)-j(x,y),k(x,y)-j(x,y))$$

$$= (k(x,y),k(x,y)) - (j(x,y),k(x,y))$$

$$- (k(x,y),j(x,y)) + (j(x,y),j(x,y))$$

$$= k(y,y) - j(y,y) - k(y,y) + j(y,y) = 0.$$

Thus $k(x,y) = j(x,y)$.

The next theorem is quite important in interpolation problems since it states that interpolation will be a well-behaved process.

Theorem 1.2. A Hilbert space X has a reproducing kernel if and only if for each fixed x, $Lf = f(x)$ is a bounded linear functional on X.

Proof. Assume that X has a reproducing kernel. Let x be fixed. Then

$$|Lf| = |f(x)| = |(f(y),k(y,x))| \leq ||f||\cdot||k(\cdot,x)||,$$

so L is a bounded linear functional.

Assume that for each fixed x, $Lf = f(x)$ is a bounded linear functional. Then by the Riesz Representation Theorem, for each y there exists a $g_y(x)$ in X such that

$$Lf = (f(x),g_y(x)) = f(y).$$

Then defining $k(x,y) = g_y(x)$, it is clear that $k(x,y)$ is a reproducing kernel on X.

Convergence in a Hilbert space usually means only norm-convergence. However, if one has a reproducing kernel, then the following property holds:

Theorem 1.3. Let X be a Hilbert space with a reproducing kernel $k(x,y)$. Let $\{f_n\}_{n=1}^{\infty}$ converge in norm to f. Then

$\{f_n\}$ converges pointwise to f and also converges uniformly to f on any set E on which k(x,x) is bounded.

Proof. Let x be fixed. Then

$$|f_n(x)-f(x)| = |(f_n(y)-f(y),k(y,x))| \leqq ||f_n-f||\cdot||k(\cdot,x)||$$

$$= ||f_n-f||\{(k(y,x),k(y,x))\}^{1/2} = ||f_n-f||\sqrt{k(x,x)}.$$

From the proof to Theorem 1.3 the following two corollaries are immediate.

Corollary 1.4. If X is a Hilbert space with a reproducing kernel k(x,y), then for any y, k(y,y) is nonnegative.

Corollary 1.5. With the hypothesis of Corollary 1.4, if for x in D there exists a function f in X such that $f(x) \neq 0$, then $k(x,x) \neq 0$.

Since most useful Hilbert spaces are separable, the following is an important theorem:

Theorem 1.6. Suppose that X is a separable Hilbert space and has a reproducing kernel k(x,y). Then if $\{\phi_n\}_{n=1}^{\infty}$ is any complete orthonormal system on X, we have

$$k(x,y) = \sum_{n=1}^{\infty} \phi_n(x)\overline{\phi_n(y)}.$$

Proof. Let $\{\phi_n\}_{n=1}^{\infty}$ be any complete orthonormal system for X. Since k(x,y) is a reproducing kernel on X, for any n,

$$(k(x,y),\phi_n(x)) = \overline{(\phi_n(x),k(x,y))} = \overline{\phi_n(y)}.$$

Thus $\overline{\phi_n(y)}$ is the Fourier coefficient of $k(x,y)$, so

$$k(x,y) = \sum_{n=1}^{\infty} \phi_n(x)\overline{\phi_n(y)} = \sum_{n=1}^{\infty} (k(x,y),\phi_n(x))\phi_n(x).$$

Corollary 1.7. If X is a finite dimensional Hilbert space, then X has a reproducing kernel, given by

$$k(x,y) = \sum_{m=1}^{n} \phi_n(x)\overline{\phi_n(y)},$$

where $\{\phi_1,\ldots,\phi_n\}$ is an orthonormal basis for X.

Since it is rather cumbersome to write the words reproducing kernel, we shall indicate them by the abbreviation r.k.. Also, in the next theorem and in later chapters, because for fixed y $k(x,y)$ is in the Hilbert space, we shall denote L operating on the function $k(x,y)$ by

$$L_x k(x,y),$$

where L is a functional on the Hilbert space.

Theorem 1.8. If L is a bounded linear functional on a Hilbert space X with r.k. $k(x,y)$, then its representer is $\overline{L_x k(x,y)}$.

Proof. Suppose g is the representer of L. Then if f is in X,

$$Lf = (f,g) = (f(x),(g(y),k(y,x)))$$

$$= (f(x),\overline{(k(y,x),g(y))}) = (f(x),\overline{L_y k(y,x)}).$$

Since by the Riesz Representation Theorem representers are unique, $g(y) = \overline{L_x k(x,y)}$.

For later use, we need the next theorem.

Theorem 1.9. Let X be a Hilbert space with a r.k. $k(x,y)$. Let $\{x_n\}_{n=1}^{M}$ be a sequence of distinct points in D such that $M \leqq \dim X$. Suppose that for each j, there exists f_j in X such that $f_j(x_j)$ is nonzero. Then $\{k(x,x_1),\ldots,k(x,x_M)\}$ is linearly independent [Note: M may be $+\infty$].

Proof. Suppose there are a positive integer n and constants $\{c_i\}_{i=1}^{n}$ such that

$$\sum_{i=1}^{n} c_i k(x,x_i) = 0.$$

Now suppose that $c_k \neq 0$. Let $N>n$, and consider the n equations

$$\delta_{ki} = \sum_{j=1}^{N} \alpha_j f_j(x_i)$$

for $i = 1,\ldots,n$. Since we have n equations in N unknowns, there will be a solution $(\alpha_1,\ldots,\alpha_N)$. But then

$$0 = \left(\sum_{j=1}^{N}\alpha_j f_j(x), \sum_{i=1}^{n} c_i k(x,x_i)\right) = \sum_{j=1}^{N}\sum_{i=1}^{n}\alpha_j \overline{c_i}(f_j(x),k(x,x_i))$$

$$= \sum_{i=1}^{n}\overline{c_i}\sum_{j=1}^{N}\alpha_j f_j(x_i) = \sum_{i=1}^{n}\overline{c_i}\delta_{ki} = \overline{c_k},$$

a contradiction.

CHAPTER II

INTERPOLATION PROBLEMS

This chapter gives some of the applications of the theory of reproducing kernels to problems involving interpolation. Section One is concerned with finding an interpolating function corresponding to a set of distinct points. A particularly useful example of a reproducing kernel Hilbert space is introduced in Section Two along with the idea of a sampling expansion. Section Three explores the uses of reproducing kernels in the theory of spline functions. An application to iterative processes is examined in Section Four.

1. Point Interpolation

In many applications the problem arises of finding that function which passes through a given point (x_o, z_o) and is of minimal norm. If the function belongs to a reproducing kernel Hilbert space X, with kernel $k(x,y)$, then the problem is solved by considering, if $k(x_o,x_o) \neq 0$,

$$\hat{f}(x) = \frac{k(x,x_o)}{k(x_o,x_o)} z_o$$

since if $f(x_o) = z_o$, then $z_o = f(x_o) = (f(x), k(x,x_o))$, and by the Cauchy-Schwartz inequality

$$|z_o| = |(f(x),k(x,x_o))| \leq ||f||\cdot||k(\cdot,x_o)||$$

$$= ||f||\sqrt{k(x_o,x_o)}$$

or

$$||f|| \geq \frac{|z_o|}{\sqrt{k(x_o,x_o)}} .$$

However,

$$\left|\left|\frac{k(\cdot,x_o)}{k(x_o,x_o)}z_o\right|\right| = \frac{|z_o|}{\sqrt{k(x_o,x_o)}}$$

implying that

$$||\hat{f}|| \leq ||f||.$$

Example 2.1. To illustrate this result, let us examine the subspace of $L_2[0,\pi]$ spanned by the orthonormal set

$$\{\sqrt{1/\pi},\ \sqrt{2/\pi}\cos x,\ \sqrt{2/\pi}\cos 2x,\ \sqrt{2/\pi}\cos 3x,\ \sqrt{2/\pi}\cos 4x\}.$$

We seek to find the function of minimal norm which passes through the point $(\pi/2,1)$. Recall from Corollary 1.7 that this subspace, since it is finite dimensional, has a reproducing kernel given by

$$k(x,y) = \sum_{p=0}^{4} g_p(x) g_p(y),$$

where $g_o(x) = \sqrt{1/\pi}$, and $g_p(x) = \sqrt{2/\pi}$ cos px, for $1 \leqq p \leqq 4$. Then the interpolate of minimal norm is

$$f(x) = \frac{k(x,\pi/2)}{k(\pi/2,\pi/2)} = \frac{(1/\pi)-(2/\pi) \cos 2x + (2/\pi) \cos 4x}{5/\pi}$$

$$= (1 - \cos 2x + \cos 4x)/5.$$

The graph of this function is illustrated in Figure 2.1.

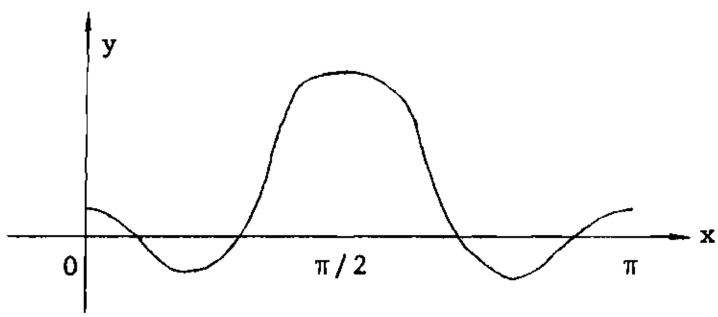

Figure 1. Interpolate of $(\pi/2,1)$.

The shape of f(x) is what one's intuition says an interpolate of minimal norm should look like: small, except for a peak at the abscissa of the interpolating point.

For the more general problem of finding a function which passes through n points, we have the following:

Theorem 2.1. Let $\{x_p\}_{p=1}^{n}$ be a sequence of distinct points, and X be a r.k. Hilbert space with r.k. k(x,y) such that for each p in $[1,n]$, there is a function g in X for which $g(x_p) \neq 0$. Then the interpolate of $\{(x_i, z_i) \mid 1 \leq i \leq n\}$ is the solution to the determinantal equation

$$(2.1) \qquad \begin{vmatrix} f(x) & z_1 & \cdots & z_n \\ k(x,x_1) & k(x_1,x_1) & \cdots & k(x_n,x_1) \\ \vdots & \vdots & & \vdots \\ k(x,x_n) & k(x_1,x_n) & \cdots & k(x_n,x_n) \end{vmatrix} = 0.$$

Moreover, f(x) is the interpolate of minimal norm.

Proof. Let M_i be the matrix obtained from

$$\begin{vmatrix} z_1 & \cdots & z_n \\ k(x_1,x_1) & \cdots & k(x_n,x_1) \\ \vdots & & \vdots \\ k(x_1,x_n) & \cdots & k(x_n,x_n) \end{vmatrix}$$

by deleting the i^{th} row. Expanding the determinant (2.1), we get

$$0 = f(x)|M_1| - k(x,x_1)|M_2|$$

$$+ k(x,x_2)|M_3| + \ldots + (-1)^n k(x,x_n)|M_{n+1}|,$$

and then solving for f,

$$f(x) = \frac{\sum_{i=1}^{n} (-1)^{n+1} k(x,x_i)\,|M_{i+1}|}{|M_1|}.$$

Notice that $|M_1| \neq 0$, since M_1 is a Gram matrix [the element $k(x_i,x_j) = (k(x,x_j),k(x,x_i))$ and by Theorem 1.9, $k(x,x_j)$ and $k(x,x_i)$ are linearly independent provided that $i \neq j$]. We can write

$$g(x) = \sum_{i=1}^{n} (-1)^{i+1} k(x,x_i)\,|M_{i+1}|$$

as

$$g(x) = (-1)^{n-1} \begin{vmatrix} z_1 & \cdots & z_n & 0 \\ k(x_1,x_1) & \cdots & k(x_n,x_1) & k(x,x_1) \\ \vdots & & \vdots & \vdots \\ k(x_1,x_n) & \cdots & k(x_n,x_n) & k(x,x_n) \end{vmatrix}.$$

Then

$$g(x_j) = (-1)^{n-1} \begin{vmatrix} z_1 & \cdots & z_n & \cdots & 0 \\ k(x_1,x_1) & \cdots & k(x_j,x_1) & \cdots & k(x_j,x_1) \\ \vdots & & \vdots & & \vdots \\ k(x_1,x_n) & \cdots & k(x_j,x_n) & \cdots & k(x_j,x_n) \end{vmatrix}$$

$$= (-1)^{n-1} \begin{vmatrix} z_1 & \cdots & z_n & -z_j \\ k(x_1,x_1) & \cdots & k(x_n,x_1) & 0 \\ \vdots & & \vdots & \\ k(x_1,x_n) & \cdots & k(x_n,x_n) & 0 \end{vmatrix}$$

$$= (-1)^{n-1}(-1)^n(-z_j)|M_1| = z_j|M_1|.$$

Thus $f(x_j) = g(x_j)/|M_1| = z_j$, so that f is an interpolate of $\{(x_i,z_i) |\ 1 \leq i \leq n\}$. Now suppose that h(x) is a function in X and $h(x_i) = f(x_i)$ for $i = 1,2,\ldots,n$. Then

$$||f||^2 = \left(\{ \sum_{i=1}^{n} (-1)^{i+1} k(x,x_i)|M_{i+1}| \}/|M_1|,\ f(x) \right)$$

$$(2.2) \qquad = \{ \sum_{i=1}^{n} (-1)^{i+1} |M_{i+1}| (k(x,x_i),f(x)) \}/|M_1|$$

$$= \{ \sum_{i=1}^{n} (-1)^{i+1} |M_{i+1}| \overline{f(x_i)} \}/|M_1|.$$

Since the value of the norm of f depends only on the values of f at the points x_i, which are equal to the values of h at each of these points, and since each step is reversible up to the first equality in (2.2), we have that

$$||f||^2 = \left(\{ \sum_{i=1}^{n} (-1)^{i+1} k(x,x_i) |M_{i+1}| \} / |M_1|, h(x) \right)$$

$$= (f(x), h(x)) \leqq ||f|| \cdot ||h||.$$

Hence f is the interpolate of minimal norm.

Recall from Chapter I that on a Hilbert space with a reproducing kernel, any linear functional of the form

$$Lf = f(x),$$

for a fixed x, is a bounded linear functional; the above theorem can therefore be generalized to interpolating n linearly independent functionals, where by interpolating a set of linear functionals $\{L^1, \ldots, L^n\}$ at the values $\{z_1, \ldots, z_n\}$, respectively, we mean finding that element f in X for which $L^i f = z_i$, for $i = 1,2,\ldots,n$.

<u>Theorem 2.2.</u> On a Hilbert space X with r.k., let L^i, $i = 1,\ldots,n$, be linearly independent bounded linear functionals. Then an interpolate of $\{(L^i, z_i) | 1 \leqq i \leqq n\}$, for z_i complex numbers, is given by

$$\begin{vmatrix} f(y) & z_1 & & z_n \\ \overline{L^1_x k(x,y)} & L^1_y \overline{L^1_x k(x,y)} & \cdots & L^n_y \overline{L^1_x k(x,y)} \\ \vdots & \vdots & & \vdots \\ \overline{L^n k(x,y)} & L^1 \overline{L^n k(x,y)} & \cdots & L^n \overline{L^n k(x,y)} \end{vmatrix} = 0.$$

Lemma 2.3. Suppose L is a bounded linear functional on a r.k. Hilbert space with r.k. $k(x,y)$. Then $L_x k(x,y)$ is an element of X.

Proof. By the Riesz Representation Theorem there is an element g in X such that $Lf = (f,g)$, for any f in X. But $L_x k(x,y) = (k(x,y), g(x)) = \overline{g(y)}$, an element of X.

Proof of Theorem 2.2. As in Theorem 2.1, suppose we express f as

$$f(x) = \frac{\sum_{i=1}^{n} (-1)^{i+1} \overline{L_x^i k(x,y)} \, |Q_{i+1}|}{|Q_{i+1}|},$$

where Q_i is the matrix

$$\begin{vmatrix} z_1 & \cdots & z_n \\ L_y^1 \overline{L_x^1 k(x,y)} & \cdots & L_y^n \overline{L_x^1 k(x,y)} \\ \vdots & & \vdots \\ L_y^1 \overline{L_x^n k(x,y)} & \cdots & L_y^n \overline{L_x^n k(x,y)} \end{vmatrix}$$

with its $i^{\underline{th}}$ row deleted. Then

$$L^j f = \{ \sum_{i=1}^{n} (-1)^{i+1} L_y^j \overline{L_x^i k(x,y)} \, |Q_{i+1}| \} / |Q_1|$$

which is the same as

$$\frac{-1}{|Q_1|}\begin{vmatrix} 0 & z_1 & \cdots & z_n \\ L_y^j\overline{L_x^1 k(x,y)} & L_y^1\overline{L_x^1 k(x,y)} & \cdots & L_y^n\overline{L_x^1 k(x,y)} \\ \vdots & \vdots & & \vdots \\ L_y^j\overline{L_x^n k(x,y)} & L_y^1\overline{L_x^n k(x,y)} & \cdots & L_y^n\overline{L_x^n k(x,y)} \end{vmatrix}$$

$$= \frac{-1}{|Q_1|}\begin{vmatrix} -z_j & z_1 \quad \cdots \quad z_n \\ 0 & \\ \vdots & Q_1 \\ 0 & \end{vmatrix} = z_j .$$

<u>Example 2.2.</u> Suppose X is the span of $p_o(x) = 1$, $p_1(x) = x$, and $p_2(x) = (3x^2 - 1)/2$, on the interval $[-1,1]$. These are Legendre polynomials, which become orthonormal polynomials on the interval $[-1,1]$ by defining

$$q_i(x) = p_i(x)\ \sqrt{2i+1}/\sqrt{2},$$

for $i = 0,1,2$, these being orthonormal functions with respect to the inner product

$$(f,g) = \int_{-1}^{1} f(x)g(x)dx.$$

Now let $L^1 f = f(1)$, and $L^2 f = \int_{-1}^{1} f(x)dx$. L^1 and L^2 are linearly independent, so we can use Theorem 2.2 to construct a function which interpolates $(L^1,-1)$ and $(L^2,0)$. By

Theorem 2.2, since

$$k(x,y) = \sum_{i=0}^{2} q_i(x)q_i(y),$$

we know that the interpolate f is given by

$$0 = \begin{vmatrix} f(y) & -1 & 0 \\ L_x^1 k(x,y) & L_y^1 L_x^1 k(x,y) & L_y^2 L_x^1 k(x,y) \\ L_x^2 k(x,y) & L_y^1 L_x^2 k(x,y) & L_y^2 L_x^2 k(x,y) \end{vmatrix},$$

or

$$f(y) = -(165y^2 + 60y - 25)/200.$$

Notice that $f(1) = -1$, and $\int_{-1}^{1} f(y)dy = 0$.

<u>Example 2.3.</u> In a more general setting, let us consider interpolation in $B = \{z \varepsilon C : ||z|| \leqq 1\}$, the unit ball in C, the set of all complex numbers. The collection X of all square-integrable analytic functions defined on B forms a Hilbert space with the inner product

$$(f,g) = \iint_B f(z)g(z)dxdy.$$

By the Cauchy Integral Theorem, if z_o is in B and S is a circle in B with z_0 in S,

$$f(z_0) = \frac{1}{2\pi i}\int_{\partial S} \frac{f(z)}{z-z_0}\, dz$$

and

$$\frac{1}{2\pi}\left|\int_{\partial S} \frac{f(z)}{z-z_0}\, dz\right| \leq \frac{1}{2\pi} \int_{\partial S} \left|\frac{f(z)}{z-z_0}\right| |dz|$$

$$\leq \frac{1}{2\pi}\,\frac{1}{\min\,|z-z_0|}\,\{\iint_S |f(z)|^2 dxdy\}^{1/2},$$

where $z = x + iy$. Thus the functional $Lf = f(z_0)$ is a bounded linear functional, implying that the Hilbert space X has a reproducing kernel. From Davis [6, p. 320] we know that $\sqrt{(n+1)/\pi}\; z^n$, $n = 0,1,2,\ldots,$ form a complete orthonormal system on B. Hence by Theorem 1.6 we can express $k(z,t)$ as

$$k(z,t) = \sum_{n=0}^{\infty} \frac{n+1}{\pi}\, z^n \bar{t}^n$$

which is the binomial expansion of

$$k(z,t) = 1/\{\pi(1-z\bar{t})^2\}.$$

Then the function which is of minimal norm and interpolates the points (0,i), (i/2,0), and (-i/2,0) is the function f which satisfies equation (2.1). This function is

$$f(t) = i[\ 544 - 225/(1-\bar{t}i/2)^2 - 225/(1+\bar{t}i/2)^2\ ]/94.$$

The following generalization of Theorem 2.2 is due to Chalmers [4].

<u>Theorem 2.3.</u> If $\{L^m\}_{m=1}^{\infty}$ is a sequence of linearly independent bounded linear functionals on a Hilbert space X with a r.k. $k(x,y)$, the problem of minimizing the norm of a function in X subject to $L^m f = a_m$, $m = 1,2,\ldots,$ has a solution if and only if $\lim_{n\to\infty} ||f_n|| = M < \infty$, where f_n is the function of minimum norm which satisfies $L^m f_n = a_m$, for $m = 1,\ldots,n$. Moreover, if a solution f exists it is unique and is given by $\lim_{n\to\infty} f_n$.

<u>Proof.</u> By Theorem 2.2 we know that f_n is the function given by

$$0 = \begin{vmatrix} f_n(y) & a_1 & \cdots & a_n \\ \overline{L_x^1 k(x,y)} & L_y^1 \overline{L_x^1 k(x,y)} & \cdots & L_y^n \overline{L_x^1 k(x,y)} \\ \vdots & \vdots & & \vdots \\ \overline{L_x^n k(x,y)} & L_y^1 \overline{L_x^n k(x,y)} & \cdots & L_y^n \overline{L_x^n k(x,y)} \end{vmatrix}.$$

Assume that $\lim_{n\to\infty} ||f_n|| = M < \infty$. Now

$$||f_n-f_m||^2 = ||f_n||^2 + ||f_m||^2 - 2\,\mathrm{Re}(f_n,f_m),$$

and if $m \geqq n$, $(f_n,f_m) = (f_n,f_n)$ since (f_n,f_m) is obtained by replacing the first column in the determinant form of f_n by

$$\begin{vmatrix} (f_n,f_m) \\ (\overline{L_x^1 k(x,y)}, f_m(y)) \\ \vdots \\ (\overline{L_x^n k(x,y)}, f_m(y)) \end{vmatrix} = \begin{vmatrix} (f_n,f_m) \\ L^1 f_m \\ \vdots \\ L^n f_m \end{vmatrix} = \begin{vmatrix} (f_n,f_m) \\ a_1 \\ \vdots \\ a_n \end{vmatrix}$$

which is almost the same column that we get when taking (f_n,f_n), except that the top element is (f_n,f_m). Thus $(f_n,f_m) = (f_n,f_n)$. Now we have

$$||f_n-f_m||^2 = ||f_n||^2 - ||f_m||^2.$$

The sequence $\{||f_n||\}$ converges, so it is Cauchy, implying that $\{f_n\}$ also forms a Cauchy sequence, which further implies that it converges to an f in X. Also,

$$L^m f = L^m(\lim_{n\to\infty} f_n) = \lim_{n\to\infty} L^m(f_n) = a_m.$$

Suppose that $L^m g = a_m$, for $m = 1,2,\ldots,$ and define $u = f-g$, so that for $m = 1,2,\ldots,$ $L^m u = 0$.

We have that

$$(u,f) = (u, \lim_{n\to\infty} f_n) = \lim_{n\to\infty} (u,f_n),$$

and examining the determinant form of f_n shows that (u,f_n) is of the form

$$\sum_{i=1}^{n} c_{in} L^i(u)$$

for some set of constants $\{c_{in} \mid 1 \leqq i \leqq n\}$. Thus

$$\lim_{n\to\infty} (u,f_n) = \lim_{n\to\infty} \sum_{i=1}^{n} c_{in} L^i(u) = 0.$$

However, (u,f) gives

$$||g||^2 = ||f||^2 + ||u||^2,$$

so that $||g|| \geqq ||f||$ and $||g|| = ||f||$ if and only if $u=0$ [or $f = g$].

Conversely, if we know that the problem has a solution, notice that

$$||f_1|| \leqq ||f_2|| \leqq \dots \leqq ||f|| < \infty,$$

so that $\lim_{n\to\infty} ||f_n|| = M < \infty$. The above arguments also show that $f = \lim_{n\to\infty} f_n$, so $||f|| = M$.

2. Sampling Expansions

In the area of electrical engineering, particularly in signal and circuit analysis, a function which represents a given signal with time as a variable is often less useful than this function transformed by a Fourier transform into the set of functions having as domain a set of frequencies. Moreover, an actual device--for example, a transmission path or an amplifier--is usually characterized as having a _finite_ bandwidth; that is, the device can only accept for processing a signal whose transformation into the frequency domain vanishes outside of a certain interval. Thus if H_1 is defined to be the subspace of $L_2(-\infty,+\infty)$ such that f is in H_1 only in case

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt$$

vanishes almost everywhere outside of $(-\pi,\pi)$, we have:

Theorem 2.4. The space H_1 is a Hilbert space with reproducing kernel

$$k(t,s) = \frac{\sin[\pi(t-s)]}{\pi(t-s)} .$$

<u>Proof.</u> If $\{f_n\}_{n=1}^{\infty}$ is a Cauchy sequence in H_1, then there is an f in $L_2(-\infty,+\infty)$ such that $\{f_n\}$ converges to f. Now we also know $\int_{-\infty}^{\infty} f_n(t)e^{-i\omega t}dt$ converges to $\int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt$, implying that this last integral also vanishes outside of $(-\pi,\pi)$. Thus H_1 is a Hilbert space. Now, if f(t) is in H_1, and $F(\omega)$ is its Fourier transform, then by the inverse Fourier transform,

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{i\omega t}d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{i\omega t}d\omega$$

so that

$$|f(t)| \leq \frac{1}{2\pi}\left[\int_{-\pi}^{\pi}|F(\omega)|^2 d\omega\right]^{1/2}\left[\int_{-\pi}^{\pi}|e^{i\omega t}|^2 d\omega\right]^{1/2}.$$

Thus the functional $Tf = f(t)$ is bounded, so that H_1 has a reproducing kernel. Now suppose that $F(\omega)$ is the Fourier transform of f. If $I(\omega)$ is the characteristic function which is one on $(-\pi,\pi)$ and zero elsewhere, and f belongs to H_1, then we have that $F(\omega) = I(\omega)F(\omega)$. Now

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} I(\omega)e^{i\omega t}d\omega = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega t}d\omega = \frac{\sin t\pi}{t\pi}.$$

Hence by the convolution theorem [7] $I(\omega)F(\omega)$ is the Fourier transform of the convolution of $f(t)$ with $(\sin \pi t)/\pi t$. Recall that the convolution of $g(t)$ and $h(t)$ is the function given by

$$c(s) = \int_{-\infty}^{\infty} g(t)h(t-s)dt.$$

Thus $F(\omega)$ is the Fourier transform of both $f(t)$ and the convolution

$$\int_{-\infty}^{\infty} \frac{f(t) \sin \pi(t-s)}{\pi(t-s)} dt,$$

so by the inverse Fourier transform,

$$f(s) = \int_{-\infty}^{\infty} f(t) \frac{\sin \pi(t-s)}{\pi(t-s)} dt.$$

Therefore the reproducing kernel is

$$k(t,s) = \frac{\sin \pi(t-s)}{\pi(t-s)}.$$

We shall, by convention, assume that the kernel has the value of one when $t = s$.

A <u>sampling expansion</u> for a set A is a type of interpolation function which depends on the value of the given function on a given set of points, the <u>sampling instants</u>. The use for a sampling expansion lies in a space of functions like H_1--those functions of finite bandwidth. In fact, Shannon [13] states that if a function f contains no frequentcies higher than W cycles/second [i.e. the Fourier transform of f vanishes for $\omega > 2\pi W$], then the function is completely determined by its values at a collection of points spaced 1/2W seconds apart, and his expansion of the function is in fact an example of a sampling expansion.

<u>Definition 2.1.</u> A class X of functions defined on a set T is said to possess a <u>sampling expansion</u> for a set of distinct <u>sampling instants</u> $\{t_i \varepsilon T \mid i \varepsilon I\}$, where I is the set of integers, if there exists a set of sampling functions

$$\{\psi_i(s,t_i) \mid i \varepsilon I\},$$

such that

(i) $\psi_i(s,t_i)$ is in X for each $i \varepsilon I$,

(ii) $\psi_i(t_j,t_i) = c_i \delta_{ij}$, $i,j \varepsilon I$, and for some set of nonzero real constants $\{c_i\}$,

(iii) for any f in X, there is a uniformly convergent expansion given by, for $s \varepsilon T$,

$$f(s) = \sum_{i \varepsilon I} f(t_i)\psi_i(s,t_i).$$

This reminds one somewhat of a complete orthonormal set, and in fact we have:

Theorem 2.5. Suppose X is a Hilbert space of functions defined on a set T, and suppose that X has a r.k. $k(s,y)$. Let

$$\{\phi_i(s,t_i) \mid t_i \varepsilon T,\ i \varepsilon I\}$$

be a complete orthonormal system in X. If there are non-zero constants $c_1, \ldots$, all real, such that

$$\phi_i(s,t_i) = c_i k(s,t_i)$$

for any $i \varepsilon I$, and

$$|k(t,t)| \leqq c < \infty,$$

for any t in T, then the complete orthonormal expansion of any f in X, given by

$$f(s) = \sum_{i \varepsilon I} a_i \phi_i(s,t_i)$$

for $s \varepsilon T$, and $a_i = (f,\phi_i)$, is a sampling expansion.

Proof. Since k is a reproducing kernel, $c_i k(s,t_i)$ is an element of X; also, $c_j \phi_i(t_j,t_i) = c_j(\phi_i(s,t_i),\ k(s,t_j))$ $= (\phi_i(s,t_i),\ \phi_j(s,t_j)) = \delta_{ij}$. Moreover, $a_i = (f,\phi_i)$

$= (f(s),\ c_i k(s,t_i)) = c_i f(t_i)$. Thus

$$f(s) = \sum_{i\varepsilon I} (f,\phi_i)\phi_i(s,t_i) = \sum_{i\varepsilon I} f(t_i)\psi_i(s,t_i),$$

where $\psi_i(s,t_i) = c_i\phi_i(s,t_i)$. We have uniform convergence because k is uniformly bounded, so in this space convergence in norm implies uniform convergence by Theorem 1.3.

In particular the set of functions

$$\left\{ \frac{\sin \pi(s-i)}{\pi(s-i)} \;\middle|\; s\varepsilon T,\ i\varepsilon I \right\}$$

forms a complete orthonormal system for H_1 and the kernel is uniformly bounded on $(-\infty,+\infty)$. Thus we have proven:

<u>Corollary 2.6.</u> Any f in H_1 possesses a sampling expansion as

$$f(s) = \sum_{i=-\infty}^{\infty} f(i)\, \frac{\sin \pi(s-i)}{\pi(s-i)}.$$

In actual usage of a sampling expansion, we would compute an approximation to the expansion, so an idea of the error involved in such an approximation is given by:

<u>Theorem 2.7.</u> If X is a Hilbert space with r.k. k(x,y), and X possesses a sampling expansion for any f in X,

$$f(s) = \sum_{I} f(t_i)\psi_i(s,t_i),$$

and if I' is a finite subset of I, then defining the truncation error to be

$$E_{I'}(s) = \sum_{I-I'} f(t_i)\psi_i(s,t_i)$$

gives

$$|E_{I'}(s)| \leqq |E - \sum_{i\epsilon I'} c_i^2 f^2(t_i)|^{1/2} \; |\sum_{I-I'} c_i^2 k^2(s,t_i)|^{1/2}$$

for any f in X" = $\{f\epsilon X : ||f||^2 \leqq E\}$, and the c_i's are as given in Theorem 2.5.

Proof. The hypercircle inequality of Golomb and Weinberger [6, p. 185] states

$$|Lf - Lf_o|^2 \leqq \left|E - ||f_o||^2\right|^{1/2} \left|\sum_{I-I'} (L\phi_k)^2\right|^{1/2},$$

where L is a bounded linear functional on a Hilbert space S, f is in S" = $\{g\epsilon S : ||g||^2 \leqq E\}$, and f_o is an element of S of minimum norm satisfying $(f_o,\phi_i) = b_i$, for i in I', with the b_i's fixed constants and $\{\phi_i \mid i\epsilon I\}$ is a complete orthonormal system for S.

Here, with $X = S$, $X'' = S''$, $b_i = f(t_i)$, then

$$f_o(s) = \sum_{i \varepsilon I'} f(t_i) c_i k(s,t_i)$$

$$= \sum_{i \varepsilon I'} f(t_i) \phi_i(s,t_i),$$

and

$$f(s) = \sum_{i \varepsilon I} f(t_i) \phi_i(s,t_i),$$

for f in X''. Also, $L_s f = (f(x),k(x,s)) = f(s)$. Thus

$$|E\quad(s)| = |L_s f - L_s f_o| = |\sum_{I-I'} f(t_i)\phi_i(s,t_i)|$$

$$\leqq |E - \sum_{I'} f^2(t_i) c_i^2|^{1/2} |\sum_{I-I'} c_i^2 k^2(s,t_i)|^{1/2},$$

which is what we had to prove.

For the specific example of Corollary 2.6, for M and N positive integers, and $i_o(s)$ the integer nearest s, let

$$I' = \{i : i_o(s) - M \leqq i \leqq i_o(s) + N\}.$$

In the context of Theorem 2.5, each $c_k = 1$, and each $t_i = i$.

Also, $\sin^2\pi(s-i) \leq 1$, so that we have

$$\left|\sum_{I-I'} c\ k\ (s,t\ )\right| = \left|\sum_{I-I'} \frac{\sin^2\pi(s-i)}{\pi^2(s-i)^2}\right| \leq \sum_{I-I'} \frac{1}{\pi^2(s-i)^2}\ ,$$

and for s in $(i_o(s), i_o(s) + 1/2]$,

$$\sum_{I-I'} \frac{1}{(s-i)^2} < \left|\frac{1}{M} + \frac{2}{2N-1}\right|,$$

while for s in $(i_o(s) - 1/2, i_o(s)]$,

$$\sum_{I-I'} \frac{1}{(s-i)^2} < \left|\frac{1}{N} + \frac{2}{2M-1}\right|,$$

so that if

$$E_{M,N}(s) = \sum_{I-I'} f(i) \frac{\sin\ \pi(s-i)}{\pi(s-i)},$$

then

$$|E_{M,N}(s)| \leq \left|E - \sum_{I'} f^2(i)\right|^{1/2} \left|\sum_{I-I'} k^2(s,i)\right|^{1/2},$$

which is in turn equal to

$$|E - \sum_{I'} f^2(i)|^{1/2} \, |\sum_{I-I'} \frac{\sin^2\pi(s-i)}{\pi^2(s-i)^2}|^{1/2}$$

$$\leqq \frac{\sqrt{E}}{\pi} \, |\sum_{I-I'} 1/(s-i)^2|^{1/2} ,$$

thus proving:

<u>Corollary 2.8.</u> For the space H_1 and sampling expansion

$$f(s) = \sum_{i=-\infty}^{\infty} f(i) \frac{\sin \pi(s-i)}{\pi(s-i)},$$

the truncation error in considering

$$\sum_{i_o(s) - M}^{i_o(s) + N} f(i) \frac{\sin \pi(s-i)}{\pi(s-i)}$$

as an approximation to $f(s)$, where N and M are positive integers, and $||f||^2 \leqq E$, is bounded by

$$|E_{M,N}(s)| < \frac{\sqrt{E}}{\pi} \begin{cases} \frac{1}{M} + \frac{2}{2N-1}, & s \text{ in } (i_o(s), i_o(s) + 1/2], \\ \frac{1}{N} + \frac{2}{2M-1}, & s \text{ in } (i_o(s) - 1/2, i_o(s)]. \end{cases}$$

3. Splines

Other types of interpolation are often used, one of the common types being piecewise linear interpolation. In fact, suppose $[a,b]$ is a finite interval and

$$a = x_o < x_1 < \ldots < x_n < x_{n+1} = b$$

is a partition of the interval. Then for $1 \leqq p \leqq n$, define

$$s_p(t) = \begin{cases} (t - x_{p-1})/(x_p - x_{p-1}), & x_{p-1} \leqq t \leqq x_p, \\ (x_{p+1} - t)/(x_{p+1} - x_p), & x_p \leqq t \leqq x_{p+1}, \\ 0, \text{ otherwise.} \end{cases}$$

Notice that $s_p(x_j) = \delta_{jp}$, so if $f(x)$ is continuous on $[a,b]$, by considering $f(x_j)s_j(t)$ we have a function which equals $f(x_j)$ when $t = x_j$ and will be zero for $t \notin (x_{j-1}, x_{j+1})$. Thus we shall interpolate with

$$\sum_{p=1}^{n} f(x_p)s_p(t).$$

If $x_j \leqq t \leqq x_{j+1}$, we have

$$\sum_{p=1}^{n} f(x_p)s_p(t) = f(x_j)s_j(t) + f(x_{j+1})s_{j+1}(t)$$

$$= f(x_j)(x_{j+1}-t)/(x_{j+1}-x_j) + f(x_{j+1})(t-x_j)/(x_{j+1}-x_j)$$

$$= \frac{f(x_{j+1}) - f(x_j)}{x_{j+1}-x_j}(t-x_j) + f(x_j).$$

It is easy to see that this last expression is the piecewise linear function passing through $(x_j, f(x_j))$ and $(x_{j+1}, f(x_{j+1}))$ on the interval $[x_j, x_{j+1}]$. The functions $\{s_p\}_{p=1}^{n}$ are called Chapeau functions, and their linear combinations, as in the interpolatory function above, are called piecewise linear splines. Now each Chapeau function is itself a continuous function with its derivative existing almost everywhere, and since $s_p(x_j) = \delta_{jp}$, we have the piecewise linear splines satisfying the first two properties of a sampling expansion, for the set $C[a,b]$. It can also be shown that over successive refinements of partitions of $[a,b]$, where these refinements in some manner approach a dense subset of $[a,b]$, that the interpolating piecewise linear splines converge uniformly to a given continuous function f. Thus we can have a uniform convergence-like property, similar to that for a sampling expansion.

In addition, these functions have several minimum properties. Suppose that f is a function which is absolutely continuous on $[a,b]$, with f' in $L_2[a,b]$ [Note: by f', the derivative of f, we shall mean that element of $L_2[a,b]$ such that $f(x) = \int_a^x f'(s)\, ds$.]. Then if Pf is the piecewise linear spline which interpolates f at each of the x_j's, then

$$\int_a^b [f'(t) - Pf'(t)]Pf'(t)\,dt$$

$$= \sum_{k=1}^{n+1} \int_{x_{k-1}}^{x_k} [f'(t) - Pf'(t)]Pf'(t)\,dt$$

$$= \sum_{k=1}^{n+1} [f(t) - Pf(t)]\Big|_{x_{k-1}}^{x_k} Pf'((x_k + x_{k-1})/2)$$

$$= 0$$

since Pf' is constant on (x_{k-1},x_k), and Pf interpolates f at each x_j. Thus we have

$$(2.4) \quad \int_a^b [f'(t)]^2dt - \int_a^b [Pf'(t)]^2dt = \int_a^b [f'(t) - Pf'(t)]^2dt$$

since

$$\int_a^b [f'(t) - Pf'(t)]^2dt = \int_a^b [f'(t)]^2dt - \int_a^b [Pf'(t)]^2dt$$

$$+ 2\int_a^b [f'(t)Pf'(t)]dt$$

$$+ 2\int_a^b [Pf'(t)]^2dt$$

$$= \int_a^b [f'(t)]^2 dt - \int_a^b [Pf'(t)]^2 dt$$

$$+ 2\int_a^b [f'(t) - Pf'(t)]Pf'(t)dt,$$

and this last term we have just shown to be zero. Hence we have

$$\int_a^b [f'(t)]^2 dt \geqq \int_a^b [Pf'(t)]^2 dt.$$

Also, in (2.4) suppose we replaced f(t) by f(t) - s(t), where s(t) is any piecewise linear spline. The interpolating function for f(t) - s(t) is Pf(t) - s(t), so that by (2.4),

$$\int_a^b [f'(t) - s'(t)]^2 dt - \int_a^b [Pf'(t) - s'(t)]^2 dt$$

$$= \int_a^b [f'(t) - s'(t) - Pf'(t) + s'(t)]^2 dt.$$

Now we have the inequality

$$\int_a^b [f'(t) - s'(t)]^2 dt \geqq \int_a^b [Pf'(t) - s'(t)]^2 dt.$$

These results do not appear especially surprising when viewed alone, but one may observe that

$$\{\int_a^b [f'(t)]^2 dt\}^{1/2}$$

is a seminorm on the space of absolutely continuous functions on the interval [a.b]. In fact, by a generalization of the above properties we can glean useful information about the process of interpolation with spline functions by a correspondence to minimum properties and best approximations. However, before pursuing that we shall need the following information:

Proposition 2.9. Let X be a Hilbert space, H be a subspace of X. Let f belong to X. Then g is the best approximation in H to f if and only if g is the orthogonal projection of f into H.

Proof. Let x be in X, and m be the best approximation to x in H. Let h belong to H. Then

$$||x-m-h||^2 = ||x-m||^2 - 2\mathrm{Re}(x-m,h) + ||h||^2$$

and since m is the best approximation to x,

$$||x-m-h|| \geq ||x-m||.$$

Thus $0 \leqq -2\mathrm{Re}(x-m,h) + ||h||^2$, or $||h||^2 \geqq 2\mathrm{Re}(x-m,h)$. I.e., for real $\lambda \neq 0$,

$$\lambda^2||h||^2 = ||h\lambda||^2 \geqq 2\lambda\mathrm{Re}(x-m,h),$$

or

$$\lambda||h||^2 \geqq 2\mathrm{Re}(x-m,h),$$

which is absurd unless $\mathrm{Re}(x-m,h) = 0$. This still true if h is replaced by ih, so that

$$(x-m,h) = \mathrm{Re}(x-m,h) - i\mathrm{Re}(x-m,-ih) = 0.$$

Therefore m is the orthogonal projection of x into H.

Notice that each step is reversible, so we have the equivalence of the orthogonal projection of x to H and the best approximation of x in H.

Now, having this fact in mind, we can generalize the piecewise linear splines given above to be more than linear functions between the x_i's by:

Definition 2.2. A polynomial spline function of degree $m \geqq 0$, having the n+1 joints $x_1 < x_2 < \ldots < x_n$ is a real valued function of class $C^{m-1}(-\infty,+\infty)$, which reduces to a polynomial of degree at most m in each of the n+1 intervals $(-\infty,x_1)$, (x_1,x_2), ..., $(x_n,+\infty)$.

Thus the piecewise linear splines, if extended to be constant outside of their original interval of definition, are polynomial splines of degree 1 with n+2 joints.

Specifically, we shall be interested in $m = 2k-1$, and $n \geq k \geq 1$. Let $[a,b]$ be a finite interval with

$$a = x_1 < \ldots < x_n = b,$$

and consider the class of real-valued functions

$$F^k[a,b] = \{f \varepsilon C^{k-1}[a,b] \mid f^{(k-1)} \text{ is absolutely continuous, and } f^{(k)} \varepsilon L_2[a,b]\}.$$

Let S be the subspace of spline functions of degree at most m which reduce to polynomials of degree at most k-1 in each of $(-\infty, x_1)$ and $(x_n, +\infty)$. Then the elements of S have the following properties [3]:

<u>Interpolation Property</u>

Given f in $F^k[a,b]$, there is a unique element Pf in S such that $Pf(x_i) = f(x_i)$, for $i = 1,\ldots,n$.

<u>First Minimum Property</u>

For f in $F^k[a,b]$, s in S, we have

$$\int_a^b [f^{(k)}(x) - s^{(k)}(x)]^2 dx \geq \int_a^b [f^{(k)}(x) - Pf^{(k)}(x)]^2 dx$$

with equality if and only if s is in $Pf + \pi_{k-1}$, where π_{k-1} are the polynomials of degree k-1 or less.

Notice that this property can be interpreted as stating that the spline interpolate is the "smoothest" approximation to f.

Second Minimum Property

If f is in $F^k[a,b]$ then

$$\int_a^b [f^{(k)}(x)]^2 dx \geq \int_a^b [Pf^{(k)}(x)]^2 dx,$$

with equality if and only if f = Pf.

As in the case of the piecewise linear splines, the above properties have been discovered independently of reproducing kernels [2]. However, the idea of the reproducing kernel gives a nice representation of these properties. First, we need a space to contain the splines, so we have the following theorem:

Theorem 2.10. The linear space $F^k[a,b]$ is a Hilbert space with respect to the inner product

$$(f,g) = \sum_{i=1}^{k} g(x_i)f(x_i) + \int_a^b f^{(k)}(y)g^{(k)}(y)dy$$

This Hilbert space possesses a reproducing kernel, for if we define

$$c_j(x) = \frac{(x-x_1)\ldots(x-x_{j-1})(x-x_{j+1})\ldots(x-x_k)}{(x_j-x_1)\ldots(x_j-x_{j-1})(x_j-x_{j+1})\ldots(x_j-x_k)}$$

then the kernel is

$$(2.5)\quad k(y,x) = \sum_{i=1}^{k} c_i(x)c_i(y) + (-1)^k\{(x-y)_+^{2k-1}$$

$$+ \sum_{i=1}^{k}\sum_{j=1}^{k}(x_i-x_j)_+^{2k-1}c_i(x)c_j(y)$$

$$- \sum_{i=1}^{k}[(x-x_i)_+^{2k-1}c_i(y) + (x_i-y)_+^{2k-1}c_i(x)]\}/(k-1)!.$$

Proof. Let $\{f_n\}_{n=1}^{\infty}$ be a Cauchy sequence of elements of $F^k[a,b]$. For each n, $f_n^{(k)}$ is an element of $L_2[a,b]$, and we have

$$||f_n-f_m||^2 = \sum_{i=1}^{k}[f_n(x_i) - f_m(x_i)]^2$$

$$+ \int_a^b [f_n^{(k)}(x) - f_m^{(k)}(x)]^2 dx$$

$$\geqq \int_a^b [f_n^{(k)}(x) - f_m^{(k)}(x)]^2 dx.$$

Thus $\{f_n\}_{n=1}^{\infty}$ is a Cauchy sequence in $L_2[a,b]$, so it converges in norm (in L_2) to a function which we shall write $f^{(k)}$. Then the function f has k arbitrary constants, which are determined by the pointwise limits of the k sequences $\{f_n(x_i)\}_{n=1}^{\infty}$ for $i = 1,\ldots,k$, which are each Cauchy since

$$||f_n - f_m||^2 \geqq \sum_{i=1}^{k} [f_n(x_i) - f_m(x_i)]^2 \geqq [f_n(x_j) - f_m(x_j)]^2$$

for any j in [1,k]. Since each of the sequences $\{f_n(x_j)\}_{n=1}^{\infty}$ converges pointwise, for $\varepsilon > 0$ there is a positive integer N so that if $n \geqq N$, and $1 \leqq j \leqq k$,

$$|f_n(x_j) - f(x_j)|^2 < \varepsilon^2/2k.$$

Also, there is an integer M such that $m \geqq M$ implies

$$\int_a^b [f_m^{(k)}(x) - f^{(k)}(x)]^2 dx < \varepsilon^2/2,$$

so that if $n \geqq \max\{N,M\}$, we have

$$||f_n - f||^2 = \sum_{i=1}^{k} [f_n(x_i) - f(x_i)]^2 + \int_a^b [f_n^{(k)}(x) - f^{(k)}(x)]^2 dx$$

$$< \sum_{i=1}^{k} \frac{\varepsilon}{2k} + \frac{\varepsilon}{2k} = \varepsilon^2 .$$

Thus $F^k[a,b]$ is a Hilbert space [since the construction is rather long, the verification that $k(y,x)$ as defined above is indeed a reproducing kernel for this space is in the appendix.].

Notice that in the summation term of the inner product that although we have n points, the use of only k points suffices, since only k constants are needed to determine a function, given its $k^{\underline{th}}$ derivative. Also, we made the stipulation that the functions are real-valued; by the obvious changes these results all work equally well for complex-valued functions.

Now let Pf denote a best approximation of an element f of $F^k[a,b]$ in the subspace S of spline functions. Since $F^k[a,b]$ is a Hilbert space, it is strictly convex [5,p.23-4], and S is a finite dimensional subspace, so the best approximation Pf in S is unique [6, p.142]. Thus, by Proposition 2.9, Pf is the orthogonal projection of f into S. Before we can prove the Interpolation Property, we need the following lemma:

Lemma 2.11. The span of $K = \{k(x,x_1),\ldots,k(x,x_n)\}$ is S.

Proof. From Greville [8, p. 3], we know that a spline function s can be written in general as

$$s(x) = p(x) + \sum_{j=1}^{n} c_j (x-x_j)_+^{2k-1}$$

for some p in π_{k-1}, where the coefficients c_i satisfy

$$\sum_{j=1}^{n} c_j x_j^r = 0, \tag{2.6}$$

for $r = 0,1,\ldots,k-1$. Since (2.6) gives us k equations in n unknowns, the dimension of its solution space is n-k. Also, π_{k-1} is of dimension k, so the dimension of S must be n-k+k or n. Now $F^k[a,b]$ is a space which is sufficiently rich to contain functions which are nonzero at each x_j, so from Chapter I we know that the set $\{k(x,x_1),\ldots,k(x,x_n)\}$ is linearly independent. It is also evident from the formula (2.5) for k(x,y) that $k(x,x_j)$ is in S for $j = 1,\ldots,n$, so that we have $S = \text{span}\{k(x,x_1),\ldots,k(x,x_n)\}$.

Theorem 2.12. Given f in $F^k[a,b]$, Pf is the unique element in S which interpolates f at each of $x_1,\ldots,x_n$.

Proof. Since an orthogonal projection is self-adjoint and P is exact on S [i.e., Ps = s for any s in S], we have

$$Pf(x_i) = (Pf(x),k(x,x_i)) = (f,Pk(x,x_i))$$

$$= (f(x),k(x,x_i)) = f(x_i),$$

and we get the uniqueness since P is single-valued.

Theorem 2.13. If f belongs to $F^k[a,b]$, and s is in S, then

$$\int_a^b [f^{(k)}(x)-s^{(k)}(x)]^2dx \geqq \int_a^b [f^{(k)}(x)-Pf^{(k)}(x)]^2dx,$$

with equality if and only if s is in $Pf + \pi_{k-1}$.

Proof. Since P is an orthogonal projection, for any g in $F^k[a,b]$ we have the Pythagorean Theorem,

$$||g||^2 = ||g-Pg||^2 + ||Pg||^2.$$

Taking g = f - s, and remembering that Ps = s for s in S,

$$||f-s||^2 = ||f-s-P(f-s)||^2 + ||P(f-s)||^2$$

$$= ||f-s-Pf+Ps||^2 + ||Pf-Ps||^2$$

$$= ||f-Pf||^2 + ||Pf-s||^2.$$

Thus

$$\sum_{i=1}^{k} [f(x_i)-s(x_i)]^2 + \int_a^b [f^{(k)}(x)-s^{(k)}(x)]^2dx$$

$$= \sum_{i=1}^{k} [f(x_i)-Pf(x_i)]^2 + \int_a^b [f^{(k)}(x)-Pf^{(k)}(x)]^2 dx$$

$$+ \sum_{i=1}^{k} [Pf(x_i)-s(x_i)]^2 + \int_a^b [Pf^{(k)}(x)-s^{(k)}(x)]^2 dx,$$

which by Theorem 2.12 is

$$\int_a^b [f^{(k)}(x)-s^{(k)}(x)]^2 dx = \int_a^b [f^{(k)}(x)-Pf^{(k)}(x)]^2 dx$$

$$+ \int_a^b [Pf^{(k)}(x)-s^{(k)}(x)]^2 dx,$$

and the last term is nonnegative, implying the First Minimum Property. The equality conclusion comes from the fact that the $k^{\underline{th}}$ derivative of s is identically zero only in case s is a polynomial of degree less than k.

The Second Minimum Property comes from the fact that, from Theorem 2.1, the interpolate of f of minimal norm is in the span of $\{k(x,x_1),\ldots,k(x,x_n)\}$, implying that in $F^k[a,b]$ such an interpolate is in S. However, by Theorem 2.12, Pf is the unique interpolate of f in S, so Pf satisfies

$$||f|| = \sum_{i=1}^{k} f^2(x_i) + \int_a^b [f^{(k)}(x)]^2 dx \geqq ||Pf||^2$$

$$= \sum_{i=1}^{k} Pf^2(x_i) + \int_a^b [Pf^{(k)}(x)]^2 dx,$$

establishing the Second Minimum Property.

4. Iteration

The following example is due to Stiefel [15], and is somewhat different from the problems we have already considered, but is an application of the results of Section One.

Suppose that X is a Hilbert space of functions defined on R, and that X contains at least the polynomials. Let A be a linear transformation from X into X such that A is an injection. We want to solve the problem

$$A(x) = k,$$

for a given k, by an iterative technique which assumes that $x_o = 0$ and defines

$$x_{i+1} = x_i + \Delta x_i$$

$$\Delta x_i = \frac{1}{q_i} r_i,$$

where q_i is called a relaxation factor and r_i is the residual

$$r_i = k - A(x_i).$$

Thus we want to choose the directions and lengths of the segments of the paths through the choice of the q_i. Notice that if A is symmetric and positive definite then we have a steepest descent method.

We are after an n-step iteration which yields the least error in some form along whatever path we choose. To obtain the error, we need the residual polynomials, which are given by

$$r_{i+1} = k - A(x_{i+1}) = k - A(x_i + \frac{1}{q_i}\Delta x_i)$$

$$= k - A(x_i + \frac{1}{q_i}r_i) = k - A(x_i) - \frac{1}{q_i}A(r_i) = r_i - \frac{1}{q_i}A(r_i)$$

$$= (I - \frac{1}{q_i}A)r_i = \ldots = (I - \frac{1}{q_i}A)\ldots(I - \frac{1}{q_i}A)k.$$

If we introduce a real variable λ, we can define

$$R_n(\lambda) = \prod_{i=0}^{n-1} (1 - \lambda/q_i).$$

Notice that $R_n(0) = 1$.

Now we introduce as a measure of the error the integral

$$\int_a^b R_n(\lambda)\,d\alpha(\lambda), \tag{2.7}$$

where $[a,b]$ contains all the roots of $R_n(\lambda)$, and where

$$(2.8) \qquad (f,g) = \int_a^b f(\lambda)g(\lambda)d\alpha(\lambda)$$

is the inner product on the Hilbert space X. We thus want to minimize (2.7), since it gives a measure of the cumulative error along the path. We have:

<u>Theorem 2.14.</u> Any polynomial R_n having degree n, real roots, and satisfying $R_n(0) = 1$ determines uniquely an iteration process yielding as last residual $r_n = R_n(A)k$, and vice versa.

This theorem is easily seen to be true since we merely take the roots of R_n to be the relaxation factors as above. In X, the polynomials of degree less than n+1 form a finite dimensional space, so they have a reproducing kernel $k_n(x,y)$. Now, from Theorem 2.1,

$$(2.9) \qquad \frac{k_n(\lambda,0)}{k_n(0,0)}$$

is the interpolate of (0,1) of minimal norm, i.e. the interpolate of (0,1) such that (2.7) is minimized. Thus by Theorem 2.14, take as relaxation factors the roots of the polynomials (2.9), which Szegö [16] has shown to be real and in $[a,b]$.

Unfortunately, this iterative process suffers from stability problems since there are n! possible paths that can be chosen, and also the roots of the polynomial (2.9) are generally quite difficult to find. We therefore make the modification which follows.

Define

$$R_i(\lambda) = \frac{k_i(\lambda,0)}{k_i(0,0)}$$

for $i = 1,\ldots,n$. The R_i's are themselves orthogonal polynomials with an inner product like (2.8) [15, p.5] provided that 0 is not in (a,b), so they will be linked by a recursive relation

$$\lambda R_i(\lambda) = -a_i R_{i+1}(\lambda) + c_i R_i(\lambda) - b_i R_{i-1}(\lambda).$$

Well, $R_i(0) = 1$, so that $c_i = a_i + b_i$, and the recursion becomes

$$\lambda R_i(\lambda) = -a_i R_{i+1}(\lambda) + (a_i + b_i) R_i(\lambda) - b_i R_{i-1}(\lambda).$$

Taking $r_o = k$, we have the residuals $r_i = R_i(A)k$, and substituting A for λ gives

$$AR_i(A) = -a_i R_{i+1}(A) + (a_i + b_i) R_i(A) - b_i R_{i-1}(A),$$

which implies

(2.10) $$Ar_i = -a_i r_{i+1} + (a_i + b_i)r \quad - b_i r_{i-1}.$$

Now if we define $\Delta r_i = r_{i+1} - r_i$ we have (2.10) as

(2.11) $$\Delta r_i = (b_i \Delta r_{i-1} - A(r_i))/a_i.$$

Now we want to construct an iteration which yields the r_i's as residuals. Again, we shall assume that $x_o = 0$, and we shall also define $\Delta x_i = x_{i+1} - x_i$, and $\Delta x_{-1} = 0$. Now,

$$\Delta r_i = r_{i+1} - r_i = k - A(x_{i+1}) - k - A(x_i) = -A(\Delta x_i),$$

so that from (2.11) we get

$$x_i = (r_i + b_i \Delta x_{i-1})/a_i.$$

Notice that $r_o = k = k - A(x_o)$, and thus by the construction

$$x_{i+1} = x_i + \Delta x_i,$$

we have designed an iterative process which yields the residuals r_i at each step, since

$$k - A(x_{i+1}) = k - A(x_i + \frac{1}{a_i} r_i + \frac{b_i}{a_i} x_{i-1})$$

$$= k - A(x_i) - \frac{1}{a_i}A(r_i) - \frac{b_i}{a_i}A(x_{i-1})$$

$$= - \frac{1}{a_i}A(r_i) + r_i + \frac{b_i}{a_i}\Delta r_{i-1},$$

which is equal to r_{i+1} since this the same expression as obtained by solving (2.10) for r_{i+1}. The advantage of this iterative scheme is that, first, it does not require knowledge of the roots of any of the R_i's. Moreover, at each step we have picked the polynomial of minimum error so that we have made each step a best step instead of trying to achieve a best path in picking all of the steps, as in the first part.

CHAPTER III

APPROXIMATION OF LINEAR FUNCTIONALS

Problems of approximation of linear functionals are our concern in Chapter III. Section One explores the basic results connecting approximation of linear functionals with approximation of their representers. The space $F^k[a,b]$ of the previous chapter is revisited in Section Two in order to examine approximation in the sense of Sard. In Section Three a constructive method of approximation is explored.

1. Approximation of Representers

Quite often when we are working with a bounded linear functional on some space, the actual computation of the values of the functional become rather difficult. For example, on the right spaces integration is a bounded linear functional, but many integrals are very difficult to compute. Thus the problem arises of approximation of the value of the integral, using perhaps

$$\sum \alpha_i L_i(f)$$

with some easily computed functionals L_i.

Suppose that we have a set $\{L_i | i=1,\ldots,n\}$ of linearly independent bounded linear functionals, and let S be the span of the set $\{\phi_1,\ldots,\phi_n\}$, where ϕ_i is the representer of

L_i. Notice that since $\{L_1,\ldots,L_n\}$ is a linearly independent set, we have $\{\phi_1,\ldots,\phi_n\}$ also linearly independent. Let L be the bounded linear functional on X which we want to approximate by a functional in the span of the L_i's in such a way that the norm of the error functional

$$R = L - \sum_{i=1}^{n} \alpha_i L_i$$

is minimized. With this background, we now have:

Theorem 3.1. Let P be the orthogonal projection of X onto a subspace S of X. If ϕ is the representer of the bounded linear functional L, then $P\phi$ is the representer of the unique best approximation $\overline{L}$ to L by a functional of the form

$$\sum_{i=1}^{n} \alpha_i L_i$$

with respect to functional norm.

Lemma 3.2. T is a bounded linear functional of the form

$$\sum_{i=1}^{n} \alpha_i L_i$$

if and only if the representer g of T is in S.

Proof. Well, T(f) is given by

$$T(f) = \sum_{i=1}^{n} \alpha_i L_i(f) = \sum_{i=1}^{n} \alpha_i (f,\phi_i) = (f, \sum_{i=1}^{n} \bar{\alpha}_i \phi_i),$$

and since by the Riesz Representation Theorem, representers are unique,

$$g = \sum_{i=1}^{n} \bar{\alpha}_i \phi_i$$

in S is the unique representer of T. Since the argument is symmetric, we have the equivalence.

Proof of Theorem 3.1. From Proposition 2.8, we know that best approximation to S and orthogonal projection to S are equivalent, so for any s in S we have

$$||\phi - s|| \geqq ||\phi - P\phi||.$$

Let $\bar{L}$ be the bounded linear functional which has $P\phi$ as its representer. By the Riesz Representation Theorem we know that $||\bar{L}|| = ||P\phi||$, the norms taken in their respective spaces. Since $P\phi$ is in the span of $\{\phi_1,\ldots,\phi_n\}$, we know by the Lemma 3.2 that $\bar{L}$ is of the form

$$\sum_{i=1}^{n} \alpha_i L_i.$$

Also, if $b_1,\ldots,b_n$ are scalars, then

$$||L - \sum_{i=1}^{n} b_i L_i|| = ||\phi - \sum_{i=1}^{n} b_i \phi_i||$$

$$= ||\phi - P\phi|| = ||L - \overline{L}||.$$

Thus $\overline{L}$ is a best approximation to L, and we have uniqueness by the fact that $P\phi$ is the unique best approximation of ϕ in S.

Corollary 3.3. With the hypotheses of Theorem 3.1, the value of the best approximation to L at f equals the value of L at the best approximation of f. I.e., $\overline{L}(f) = L(Pf)$.

Proof. Well, $\overline{L}(f) = (f,P\phi) = (Pf,\phi) = L(Pf)$, since an orthogonal projection is self-adjoint.

Corollary 3.4. The best approximation $\overline{L}$ to L is exact on S.

Proof. If s is in S, we have that $Ps = s$, so that

$$\overline{L}(s) = L(Ps) = L(s).$$

Example 3.1. Let us return to the space $F^k[a,b]$, and consider S as the set of spline functions we had before. We shall define the linear functionals $L_i(f) = f(x_i)$, for $i=1,\ldots,n$. Since $F^k[a,b]$ is a r.k. Hilbert space we know that each L_i is bounded. Also, the representer of L_i is $k(x,x_i)$ since $(f(x),k(x,x_i)) = f(x_i)$. Recall from Lemma 2.11 we know that $S = \text{span}\{k(x,x_1),\ldots,k(x,x_n)\}$. Now let L be a bounded linear functional on $F^k[a,b]$ and suppose we want to find the best approximation $\overline{L}$ to L of the form

$$\sum_{i=1}^{n} \alpha_i f(x_i).$$

We want the best approximation in the same sense as in Theorem 3.1--that of minimizing the norm of the residual. Thus if ϕ is the representer of L, then the representer of $\overline{L}$ is $P\phi$, which is the spline interpolate of ϕ.

To be specific, consider the problem of integration on the interval [0,1], in the space $F^1[0,1]$. The linear functional L is thus

$$L(f) = \int_0^1 f(x)dx.$$

Let the number of joints be three, namely $x_1 = 0$, $x_2 = 1/3$, and $x_3 = 2/3$. Then the functionals L_i are defined by $L_i(f) = f((i-1)/3)$, and the representer of L is [In the formula for the r.k., we assume $c_1(x) = 1$ when k=1.],

$$L_x k(x,y) = 1 + y - y^2/2,$$

and the reproducing kernel is

$$k(x,y) = 1 + y - (y-x)_+ = \begin{cases} 1+x, & y \geqq x \\ 1+y, & x \geqq y \end{cases}$$

for the inner product taken with respect to x, when the inner product is

$$(f,g) = f(0)g(0) + \int_0^1 f'(x)g'(x)dx.$$

From Chapter II, we know that the best approximation $P(L_x k(x,y))$ is merely the interpolate in the space of spline functions, which is given by

$$s(x) = P(L_x k(x,y)) = 1 + \frac{5}{6}x - \frac{1}{3}(x-1/3)_+ - \frac{1}{2}(x-2/3)_+.$$

Then if f is in $F^1[0,1]$, we have

$$(f(x),s(x)) = f(0) + \int_0^1 f'(x)\frac{5}{6}\,dx$$

$$- \int_0^1 \frac{1}{3}f'(x)\{\begin{matrix} 1, x \geqq 1/3 \\ 0, x < 1/3 \end{matrix}\}dx - \int_0^1 \frac{1}{2}f'(x)\{\begin{matrix} 1, x \geqq 2/3 \\ 0, x < 2/3 \end{matrix}\}dx$$

$$= f(0)/6 + f(1/3)/3 + f(2/3)/2$$

$$= L_1(f)/6 + L_2(f)/3 + L_3(f)/2.$$

Comparing the answers given by this method, for $f(x) = x^2$, we observe that $(x^2, s(x)) = 5/18$, as compared to an exact value of $1/3$, and for $f(x) = c$, a constant,

$$(c, s(x)) = c/6 + c/3 + c/2 = c,$$

the exact result which was guaranteed by Corollary 3.4.

2. Approximation in the Sense of Sard

Now let us consider, in the space $F^k[a,b]$, the best approximation to a bounded linear functional L in the sense of Sard [3]. Since $n \geqq k$, there are numbers $a_1, \ldots, a_n$, such that

$$R(f) = L(f) - \sum_{i=1}^{n} a_i L_i(f) = 0$$

for any polynomial of degree less than k-1. Let Q be the set of functionals T of the form

$$T = \sum_{i=1}^{n} c_i L_i$$

which annihilate all polynomials of degree less than k-1. By Peano's Theorem the residual R is

$$R(g) = \int_a^b g^{(k)}(t)K(t)dt$$

for any g in $F^k[a,b]$, where K is the Peano kernel

$$K(t) = R_x[(x-t)_+^{k-1}]/(k-1)!.$$

Definition 3.1. L*, the best approximation to L in the sense of Sard, is the bounded linear functional in Q which minimizes

$$(3.1) \qquad \int_a^b K^2(t)dt.$$

For L' in Q we know that

$$||R|| = ||L-L'|| = \sup_{||f||\leq 1} |R(f)|$$

$$= \sup_{||f||\leq 1} \left|\int_a^b K(t)f^{(k)}(t)dt\right|$$

$$\leq \sup_{||f||\leq 1} \left|\int_a^b K^2(t)dt\right|^{1/2} \left|\int_a^b (f^{(k)}(t))^2 dt\right|^{1/2}$$

$$\leq \left|\int_a^b K^2(t)dt\right|^{1/2} .$$

Also, if

$$f(t) = \int_a^t \int_a^{t_1} \cdots \int_a^{t_{k-1}} \frac{K(t_k)}{C} dt_k dt_{k-1} \cdots dt_1 ,$$

where

$$C = \left|\int_a^b K^2(t)\ dt\right|^{1/2} ,$$

then $f^{(k)}(t) = K(t)/C$, and

$$||R|| \geq \left|\int_a^b K(t)f^{(k)}(t)dt\right|$$

$$= \left|\int_a^b K(t)\frac{K(t)}{C}dt\right| = \left|\int_a^b K^2(t)dt\right|^{1/2} .$$

Thus minimizing the integral (3.1) also minimizes the norm of L-L', so we have $\overline{L} = L^*$ since by Corollary 3.4 $\overline{L}$ is exact on S, implying that $\overline{L}$ - L annihilates the polynomials of degree less than k-1, so that $\overline{L}$ is in Q. We thus know that best approximation in operator norm to the set of operators of the form

$$\sum \alpha_i L_i$$

is also the best approximation in the sense of Sard in the space $F^k[a,b]$.

It is also of interest to consider the error

$$R(f) = L(f) - \overline{L}(f).$$

Since $L(f) - \overline{L}(f) = (f,\phi-P\phi) = ((I-P)f,\phi)$, by the Cauchy-Schwartz inequality we have

$$|L(f) - \overline{L}(f)| \leqq ||f||\cdot||\phi-P\phi||,$$

and

$$|L(f) - \overline{L}(f)| \leqq ||\phi||\cdot||f-Pf||.$$

However, P is an orthogonal projection, so I-P is also an orthogonal projection, implying that

$$(f,(I-P)\phi) = (f,(I-P)^2\phi) = ((I-P)f,(I-P)\phi),$$

so that we have another estimate

$$|L(f) - \bar{L}(f)| \leqq ||f-Pf||\cdot||\phi-P\phi||.$$

Then if $||f|| \leqq r$, since in Chapter II we found that

$$||f||^2 = ||f-Pf||^2 + ||Pf||^2,$$

and we have

$$||f-Pf||^2 \leqq (r^2 - ||Pf||^2)$$

so that

$$|L(f) - \bar{L}(f)| \leqq ||\phi-P\phi||\sqrt{r^2 - ||Pf||^2}.$$

Notice that now we have a bound which depends on f only to the extent of Pf; i.e., if $\bar{S}$ is the orthogonal complement of S, then we have the same bound for any g in Pf + $\bar{S}$ for which $||g|| \leqq r$.

3. Constructive Techniques

In Section One we showed that an approximation of a functional L by a linear combination of the functionals L_i, for $1 \leqq i \leqq n$, treated the span of the representers of the functionals exactly. Then in the space $F^k[a,b]$ we know that for splines with joints $x_1,...,x_n$, $k(x,x_i)$ is in the span of the representers $k(x,x_1),...,k(x,x_n)$ so $k(x,x_i)$ is treated exactly, for each i in [1,n]. It is now of interest to see a proof showing how the interpolating constants are chosen, so while the following theorem is actually a corollary to Corollary 3.4, we shall demonstrate a constructive proof.

Theorem 3.5. For a bounded linear functional L on a r.k. Hilbert space, the best approximation L' of the form

$$\sum \alpha_i f(x_i)$$

treats the functions $k(x,x_1),\ldots,k(x,x_n)$exactly, for fixed distinct $x_1,\ldots,x_n$.

Proof. As before, if $\phi(x)$ is the representer of L, and

$$R(f) = L(f) - \sum_{i=1}^{n} \alpha_i f(x_i)$$

then

$$R(f) = (f,\phi) - \sum_{i=1}^{n} \alpha_i (f(x),k(x,x_i))$$

$$= (f(x),\phi(x) - \sum_{i=1}^{n} \overline{\alpha}_i k(x,x_i)),$$

and by the Cauchy-Schwartz inequality,

$$|R(f)|^2 \leq ||f||^2 ||\phi - \sum_{i=1}^{n} \overline{\alpha}_i k(\cdot, x_i)||^2,$$

so we shall be interested in minimizing the positive definite quadratic form

$$S = ||\phi - \sum_{i=1}^{n} \overline{\alpha}_i k(\cdot, x_i)||^2.$$

Using the calculus of variations, if α_k varies a small amount $\delta\alpha_k$, we have

$$(\delta S)_{\alpha_k} = (-\delta\overline{\alpha}_k k(x, x_k), \phi(x) - \sum_{i=1}^{n} \overline{\alpha}_i k(x, x_i))$$

$$+ (\phi(x) - \sum_{i=1}^{n} \overline{\alpha}_i k(x, x_i), -\delta\overline{\alpha}_k k(x, x_k))$$

$$= -\delta\overline{\alpha}_k \left(\overline{\phi(x_k)} - \sum_{i=1}^{n} \alpha_i \overline{k(x_k, x_i)}\right)$$

$$-\delta\alpha_k \left(\varphi(x) - \sum_{i=1}^{n} \overline{\alpha}_i k(x_k, x_i)\right)$$

Thus $(\delta S)_{\alpha_k}$ vanishes for each k, if

$$\phi(x_k) = \sum_{i=1}^{n} \overline{\alpha}_i k(x_k, x_i).$$

Since $k(x_k,x_i) = (k(x,x_i),k(x,x_i))$ and the x_i's are distinct, as in the proof to Theorem 2.1, the matrix of this linear system is a Gram matrix [assuming that the space is rich enough to have a nonzero function at each x_i] and is hence nonsingular, so there will be a solution. Since the representer of the residual is as given above, for optimal choice of the α_i's,

$$R(k(x,x_k)) = (k(x,x_k), \phi(x) - \sum_{i=1}^{n} \overline{\alpha}_i k(x,x_i))$$

$$= \overline{\phi(x_k)} - \sum_{i=1}^{n} \alpha_i \overline{k(x_k,x_i)} = 0.$$

Now, instead of fixing the abscissae and choosing the weights, we shall examine the case in which the weights are fixed but we may vary the abscissae.

<u>Theorem 3.6.</u> In a r.k. Hilbert space, if the weights w_i are prescribed, then the best approximation of the form

$$\sum_{i=1}^{n} w_i f(x_i)$$

to a bounded linear functional L has the property that it treats the functions

$$\frac{\partial}{\partial x_k} k(x, x_k),$$

for $k = 1, \ldots, n$, exactly, provided that the derivatives of k, and derivatives of the representer of L exist and are in X.

Proof. If $\phi(x)$ is the representer of L, and

$$R(f) = L(f) - \sum_{i=1}^{n} w_i f(x_i),$$

then we have the same positive definite quadratic form to minimize as in Theorem 3.5,

$$S = ||\phi - \sum_{i=1}^{n} \overline{w}_i k(\cdot, x_i)||^2.$$

Here, if we make a small change δx_k in x_k we have, provided the derivatives exist,

$$(\delta S)_{x_k} = \left(\overline{\frac{d\phi}{dx_k}(x_k)} - \sum_{i=1}^{n} w_i \frac{\partial}{\partial x} \overline{k(x_k, x_i)}\right) \overline{\delta x_k}$$

$$+ \left(\frac{d\phi}{dx_k}(x_k) - \sum_{i=1}^{n} \overline{w}_i \frac{\partial}{\partial x_k} k(x_k, x_i)\right) \delta x_k .$$

Thus this expression vanishes for

$$\frac{d}{dx_k}(x_k) = \sum_{i=1}^{n} \overline{w}_i \frac{\partial}{\partial x_k} k(x_k, x_i).$$

Now if we have a choice of x_i's that minimizes S, $(\delta S)_{x_k}$ will be zero for each x_k. Notice that

$$\left(\frac{\partial}{\partial x_k} k(x, x_k), \phi(x)\right) = \left(\lim_{h \to 0} \frac{k(x, x_k + h) - k(x, x_k)}{h}, \phi(x)\right)$$

$$= \lim_{h \to 0} \left(\frac{k(x, x_k + h) - k(x, x_k)}{h}, \phi(x)\right)$$

$$= \lim_{h \to 0} \frac{1}{h} \left[\overline{\phi(x_k + h)} - \overline{\phi(x_k)}\right]$$

$$= \overline{\phi'(x_k)}.$$

Thus

$$R\left(\frac{\partial}{\partial x_k} k(x, x_k)\right) = \left(\frac{\partial}{\partial x_k} k(x, x_k), \phi(x) - \sum_{i=1}^{n} \overline{w}_i k(x, x_i)\right)$$

$$= \phi'(x_k) - \sum_{i=1}^{n} w_i \frac{\partial}{\partial x_k} k(x_i, x_k) = 0.$$

If we leave weights and abscissae free, we have:

Corollary 3.7. If the appropriate derivatives and distinct points exist, then an approximation of the form

$$\sum_{i=1}^{n} w_i f(x_i)$$

with neither the weights w_i nor the points x_i prescribed, is exact for the 2n functions

$$\{k(x,x_i), \frac{\partial}{\partial x_i} k(x,x_i) \mid i = 1,\ldots,n\}.$$

Unfortunately, the existence of the appropriate derivatives is not a trivial limitation, as the following simple example shows.

Example 3.2. Let X be the subspace of $F^1[0,1]$ consisting of those functions which vanish at the origin. This space is a Hilbert space with the inner product defined by

$$(f,g) = \int_0^1 f'(x)g'(x)dx.$$

The kernel function for this space is

$$k(x,y) = \begin{cases} x, & y \geq x \\ y, & x \geq y. \end{cases}$$

Notice that

$$\frac{\partial}{\partial y} k(x,y) = \begin{cases} 0, & y \geq x \\ 1, & x \geq y \end{cases}$$

is undefined for x = y, which is bad since the equation defining the points in the proof of Theorem 3.6 becomes meaningless for $x_i = x_k$. Thus even in the relatively simple example given here, we cannot use Theorem 3.6.

Example 3.3. Let X be the Hilbert space of functions of a complex variable x which are analytic inside the region $|x| < r$, and continuous for $|x| = r$, where r is a positive number. The inner product is defined as

$$(f,g) = \int_{|x|=r} f(x)\overline{g(x)}\ |dx|$$

and the reproducing kernel is the Szegö kernel [9]

$$k(x,y) = r/[2\pi(r^2 - y\bar{x})].$$

From Corollary 3.7 we know that a best approximation of the form

$$\sum \alpha_i f(x_i)$$

to a bounded linear functional must be exact for the 2n functions

$$\frac{r}{2\pi}(r^2 - y\bar{x}_k)^{-1}, \ \frac{ry}{2\pi}(r^2 - y\bar{x}_k)^{-2},$$

for $k = 1,\ldots,n$, and the best approximation will entail choosing the α_i's and the x_i's.

Now it will in general be possible to express the 2n functions

$$q_j(y) = y^j \prod_{k=1}^{n} (1 - \frac{y\bar{x}_k}{r^2})^{-2}$$

for $j = 0,1,\ldots,2n-1$, as a linear combination of the above functions, so that the $q_j(y)$ must also be treated exactly by the best approximation to a given functional. However, as $r \to \infty$,

$$q_j(y) \to y^j,$$

for each j, so these limiting functions achieve exact treatment in the limit as r becomes infinite.

Thus if we consider Gaussian quadrature, i.e. approximate integration of the form

$$\sum A_i f(x_i)$$

we know that for optimal choice of the A_i's, the Gaussian quadrature will be exact for polynomials of degree less than $2n$. Hence in this space, Gaussian quadrature appears as the limit as $r \to \infty$ of a sequence of optimal quadrature rules.

APPENDIX

To show that $k(y,x)$ as defined in Theorem 2.10 is a r.k. on $F^k[a,b]$. For convenience, we shall divide k into five parts,

$$p_1 = \sum_{j=1}^{k} c_j(x)c_j(y),$$

$$p_2 = (x-y)_+^{2k-1}$$

$$p_3 = \sum_{m=1}^{k} \sum_{j=1}^{k} (x_m-x_j)_+^{2k-1} c_m(x)c_j(y),$$

$$p_4 = -\sum_{j=1}^{k} (x-x_j)_+^{2k-1} c_j(y),$$

and

$$p_5 = -\sum_{j=1}^{k} (x_j-y)_+^{2k-1} c_j(x).$$

Well,

$$(f(y),p_1) = (f(y), \sum_{j=1}^{k} c_j(x)c_j(y))$$

$$= \sum_{i=1}^{k} f(x_i) \sum_{j=1}^{k} c_j(x_i) c_j(x) + \int_a^b f^{(k)}(y) \sum_{j=1}^{k} c_j(x) c_j^{(k)}(y)\,dy$$

$$= \sum_{i=1}^{k} f(x_i) c_i(x).$$

Next,

$$(f(y), p_2) = \sum_{i=1}^{k} f(x_i)(x-x_i)_+^{2k-1} + \int_a^b f^{(k)}(y)[(x-y)_+^{2k-1}]^{(k)}\,dy$$

$$= \sum_{i=1}^{k} f(x_i)(x-x_i)_+^{2k-1} + \int_a^x f^{(k)}(y)(-1)^k \frac{(2k-1)!}{(k-1)!}(x-y)^{k-1}\,dy$$

$$= \sum_{i=1}^{k} f(x_i)(x-x_i)_+^{2k-1} + (-1)^k (2k-1)! f(x)$$

$$- (-1)^k (2k-1)! \sum_{i=0}^{k-1} \frac{f^{(i)}(a)}{i!} (x-a)^i.$$

Also,

$$(f(y), p_3) = \sum_{i=1}^{k} f(x_i) \sum_{m=1}^{k} \sum_{j=1}^{k} (x_m - x_j)_+^{2k-1} c_m(x) c_j(x_i)$$

$$+ \int_a^b f^{(k)}(y) \sum_{m=1}^{k} \sum_{j=1}^{k} (x_m - x_j)_+^{2k-1} c_m(x) c_j^{(k)}(y)\,dy$$

$$= \sum_{i=1}^{k} f(x_i) \sum_{m=1}^{k} (x_m - x_i)_+^{2k-1} c_m(x).$$

Now

$$(f(y), p_4) = -\sum_{i=1}^{k} f(x_i) \sum_{j=1}^{k} (x - x_j)_+^{2k-1} c_j(x_i)$$

$$- \int_a^b f^{(k)}(y) \sum_{j=1}^{k} (x - x_j)_+^{2k-1} c_j^{(k)}(y)\,dy$$

$$= -\sum_{i=1}^{k} f(x_i)(x - x_i)_+^{2k-1}.$$

Lastly,

$$(f(y), p_5) = -\sum_{i=1}^{k} f(x_i) \sum_{j=1}^{k} (x \;\; - x_i)_+^{2k-1} c_j(x)$$

$$-\int_a^b f^{(k)}(y) \sum_{i=1}^{k} [(x_i - y)_+^{2k-1}]^{(k)} c_i(x)\,dy$$

$$= - \sum_{i=1}^{k} f(x_i) \sum_{j=1}^{k} (x_j - x_i)_+^{2k-1} c_j(x)$$

$$- \sum_{j=1}^{k} c_j(x)(-1)^k \frac{(2k-1)!}{(k-1)!} \int_a^{x_j} (x_j - y)_+^{k-1} f^{(k)}(y)\,dy$$

$$= - \sum_{i=1}^{k} f(x_i) \sum_{j=1}^{k} (x_j - x_i)_+^{2k-1} c_j(x)$$

$$- \sum_{j=1}^{k} c_j(x)(-1)^k (2k-1)! \{ f(x_j) - \sum_{i=0}^{k-1} \frac{f^{(i)}(a)}{i!} (x_j - a)^i \}.$$

Combining the above,

$$(f(y), k(y,x)) = \frac{(-1)^k}{(2k-1)!}(f(y), p_4) + \frac{(-1)^k}{(2k-1)!}(f(y), p_2)$$

$$+ \frac{(-1)^k}{(2k-1)!}(f(y), p_3) + (f(y), p_1) + \frac{(-1)^k}{(2k-1)!}(f(y), p_5)$$

$$= - \frac{(-1)^k}{(2k-1)!} \sum_{i=1}^{k} f(x_i)(x - x_i)_+^{2k-1} + \frac{(-1)^k}{(2k-1)!} \sum_{i=1}^{k} f(x_i)(x - x_i)_+^{2k-1}$$

$$+ f(x) - \sum_{i=0}^{k-1} \frac{f^{(i)}(a)}{i!} (x-a)^i$$

$$+ \frac{(-1)^k}{(2k-1)!} \sum_{i=1}^{k} f(x_i) \sum_{m=1}^{k} (x_m - x_i)_+^{2k-1} c_m(x) + \sum_{i=1}^{k} f(x_i) c_i(x)$$

$$- \frac{(-1)^k}{(2k-1)!} \sum_{i=1}^{k} f(x_i) \sum_{j=1}^{k} (x_j - x_i)_+^{2k-1} c_j(x)$$

$$- \sum_{j=1}^{k} f(x_j) c_j(x) + \sum_{j=1}^{k} c_j(x) \sum_{i=0}^{k-1} \frac{f^{(i)}(a)}{i!}(x_j - a)^i$$

$$= f(x) - \sum_{i=0}^{k-1} \frac{f^{(i)}(a)}{i!}(x-a)^i + \sum_{j=1}^{k} c_j(x) \sum_{i=0}^{k-1} \frac{f^{(i)}(a)}{i!}(x_j - a)^i$$

$$= f(x) - \sum_{i=0}^{k-1} \frac{f^{(i)}(a)}{i!} \{(x-a)^i - \sum_{j=1}^{k} c_j(x)(x_j - a)^i\}.$$

The last term is the Lagrange interpolating polynomial for $(x-a)^i$, which is equal to $(x-a)^i$, and the expression in the braces is therefore zero. Thus we have

$$f(x) = (f(y), k(y,x)).$$

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] N. Aronszajn, "Theory of Reproducing Kernels," Transactions of the American Mathematics Society, 1950, Volume 68, pages 337-404.

[2] C. de Boor, "Best Approximation Properties of Spline Functions of Odd Degree," Journal of Mathematics and Mechanics, 1963, Volume 12, pages 747-749.

[3] C. de Boor and R.E. Lynch, "On Splines and Their Minimum Properties," Journal of Mathematics and Mechanics, 1966, Volume 15, pages 953-969.

[4] B. Chalmers, "Subspace Kernels and Minimum Problems in Hilbert Spaces with Reproducing Kernels," Pacific Journal of Mathematics, 1969, Volume 31, pages 619-628.

[5] E. Cheney, Introduction to Approximation Theory, McGraw-Hill, New York, 1966.

[6] P. Davis, Interpolation and Approximation, Blaisdell, New York, 1963.

[7] P. Goldberg, Fourier Transforms, Cambridge University Press, Cambridge, 1961.

[8] T. Greville, Theory and Applications of Spline Functions, Academic Press, New York, 1969.

[9] F. Larkin, "Optimal Approximation in Hilbert Spaces with Reproducing Kernel Functions," Mathematics of Computation, 1970, Volume 24, pages 911-921.

[10] H. Meschkowski, Hilbertsche Raume mit Kernfunktion, Springer-Verlag, Berlin, 1962.

[11] I. Peterson, "Recovery of a Function in the Class of Functions with a Reproducing Kernel," Automation and Remote Control, June 1969, Volume 6, pages 918-921.

[12] A. Sard, Linear Approximations, American Mathematics Society, New York, 1963.

[13] C. Shannon, "Communication in the Presence of Noise," Proceedings of the I.R.E., January, 1949, pages 10-21.

[14] J. Singer, Elements of Numerical Analysis, Academic Press, New York, 1964.

[15] E. Stiefel, "Kernel Polynomials in Linear Algebra and Their Numerical Applications," National Bureau of Standards Applied Mathematics Series, 1958, Volume 49, pages 1-22.

[16] Szego, "Uber Orthogonale Polynome, die zy einer gegebenen Kurve der Komplexen Ebene, Gehoren," Mathematische Zeitschrift, 1921, Volume 9, pages 241-244.

[17] K. Yao, "Applications of Reproducing Kernel Hilbert Spaces-Bandlimited Signal Models," Information and Control, 1967, Volume 11, pages 429-444.

[18] K. Yosida, Functional Analysis, Springer-Verlag, Berlin, 1962.